SEQUENCE LISTING

(1) GENERAL INFORMATION:

   (i) APPLICANT: Mark Gijzen

   (ii) TITLE OF INVENTION: Seed Coat Specific DNA Regulatory Region And Peroxidase

   (iii) NUMBER OF SEQUENCES: 19

   (iv) COMPUTER READABLE FORM:
      (A) MEDIUM TYPE: Floppy disk
      (B) COMPUTER: IBM PC compatible
      (C) OPERATING SYSTEM: PC-DOS/MS-DOS
      (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

   (vi) CURRENT APPLICATION DATA:
      (A) APPLICATION NUMBER: 08/939,905
      (B) FILING DATE: 30-Sept-1996

   (vii) PRIOR APPLICATION DATA:
      (A) APPLICATION NUMBER: 08/723,414
      (B) FILING DATE: 30-Sept-1996

(2) INFORMATION FOR SEQ ID NO: 1:

   (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 1244 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: cDNA

   (iii) HYPOTHETICAL: NO

   (iv) ANTI-SENSE: NO

   (ix) FEATURE:
      (A) NAME/KEY: CDS
      (B) LOCATION:1..1056

   (ix) FEATURE:
      (A) NAME/KEY: sig_peptide
      (B) LOCATION:1..78

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
ATG GGT TCC ATG CGT CTA TTA GTA GTG GCA TTG TTG TGT GCA TTT GCT      48
Met Gly Ser Met Arg Leu Leu Val Val Ala Leu Leu Cys Ala Phe Ala
 1               5                  10                  15

ATG CAT GCA GGT TTT TCA GTC TCT TAT GCT CAG CTT ACT CCT ACG TTC      96
Met His Ala Gly Phe Ser Val Ser Tyr Ala Gln Leu Thr Pro Thr Phe
             20                  25                  30

TAC AGA GAA ACA TGT CCA AAT CTG TTC CCT ATT GTG TTT GGA GTA ATC     144
Tyr Arg Glu Thr Cys Pro Asn Leu Phe Pro Ile Val Phe Gly Val Ile
         35                  40                  45
```

```
TTC GAT GCT TCT TTC ACC GAT CCC CGA ATC GGG GCC AGT CTC ATG AGG        192
Phe Asp Ala Ser Phe Thr Asp Pro Arg Ile Gly Ala Ser Leu Met Arg
 50                  55                  60

CTT CAT TTT CAT GAT TGC TTT GTT CAA GGT TGT GAT GGA TCA GTT TTG        240
Leu His Phe His Asp Cys Phe Val Gln Gly Cys Asp Gly Ser Val Leu
 65                  70                  75                  80

CTG AAC AAC ACT GAT ACA ATA GAA AGC GAG CAA GAT GCA CTT CCA AAT        288
Leu Asn Asn Thr Asp Thr Ile Glu Ser Glu Gln Asp Ala Leu Pro Asn
                 85                  90                  95

ATC AAC TCA ATA AGA GGA TTG GAC GTT GTC AAT GAC ATC AAG ACA GCG        336
Ile Asn Ser Ile Arg Gly Leu Asp Val Val Asn Asp Ile Lys Thr Ala
                100                 105                 110

GTG GAA AAT AGT TGT CCA GAC ACA GTT TCT TGT GCT GAT ATT CTT GCT        384
Val Glu Asn Ser Cys Pro Asp Thr Val Ser Cys Ala Asp Ile Leu Ala
                115                 120                 125

ATT GCA GCT GAA ATA GCT TCT GTT CTG GGA GGA GGT CCA GGA TGG CCA        432
Ile Ala Ala Glu Ile Ala Ser Val Leu Gly Gly Gly Pro Gly Trp Pro
 130                 135                 140

GTT CCA TTA GGA AGA AGG GAC AGC TTA ACA GCA AAC CGA ACC CTT GCA        480
Val Pro Leu Gly Arg Arg Asp Ser Leu Thr Ala Asn Arg Thr Leu Ala
145                 150                 155                 160

AAT CAA AAC CTT CCA GCA CCT TTC TTC AAC CTC ACT CAA CTT AAA GCT        528
Asn Gln Asn Leu Pro Ala Pro Phe Phe Asn Leu Thr Gln Leu Lys Ala
                165                 170                 175

TCC TTT GCT GTT CAA GGT CTC AAC ACC CTT GAT TTA GTT ACA CTC TCA        576
Ser Phe Ala Val Gln Gly Leu Asn Thr Leu Asp Leu Val Thr Leu Ser
                180                 185                 190

GGT GGT CAT ACG TTT GGA AGA GCT CGG TGC AGT ACA TTC ATA AAC CGA        624
Gly Gly His Thr Phe Gly Arg Ala Arg Cys Ser Thr Phe Ile Asn Arg
                195                 200                 205

TTA TAC AAC TTC AGC AAC ACT GGA AAC CCT GAT CCA ACT CTG AAC ACA        672
Leu Tyr Asn Phe Ser Asn Thr Gly Asn Pro Asp Pro Thr Leu Asn Thr
 210                 215                 220

ACA TAC TTA GAA GTA TTG CGT GCA AGA TGC CCC CAG AAT GCA ACT GGG        720
Thr Tyr Leu Glu Val Leu Arg Ala Arg Cys Pro Gln Asn Ala Thr Gly
225                 230                 235                 240

GAT AAC CTC ACC AAT TTG GAC CTG AGC ACA CCT GAT CAA TTT GAC AAC        768
Asp Asn Leu Thr Asn Leu Asp Leu Ser Thr Pro Asp Gln Phe Asp Asn
                245                 250                 255

AGA TAC TAC TCC AAT CTT CTG CAG CTC AAT GGC TTA CTT CAG AGT GAC        816
Arg Tyr Tyr Ser Asn Leu Leu Gln Leu Asn Gly Leu Leu Gln Ser Asp
                260                 265                 270

CAA GAA CTT TTC TCC ACT CCT GGT GCT GAT ACC ATT CCC ATT GTC AAT        864
Gln Glu Leu Phe Ser Thr Pro Gly Ala Asp Thr Ile Pro Ile Val Asn
                275                 280                 285

AGC TTC AGC AGT AAC CAG AAT ACT TTC TTT TCC AAC TTT AGA GTT TCA        912
Ser Phe Ser Ser Asn Gln Asn Thr Phe Phe Ser Asn Phe Arg Val Ser
 290                 295                 300
```

```
ATG ATA AAA ATG GGT AAT ATT GGA GTG CTG ACT GGG GAT GAA GGA GAA        960
Met Ile Lys Met Gly Asn Ile Gly Val Leu Thr Gly Asp Glu Gly Glu
305             310             315             320

ATT CGC TTG CAA TGT AAT TTT GTG AAT GGA GAC TCG TTT GGA TTA GCT       1008
Ile Arg Leu Gln Cys Asn Phe Val Asn Gly Asp Ser Phe Gly Leu Ala
            325             330             335

AGT GTG GCG TCC AAA GAT GCT AAA CAA AAG CTT GTT GCT CAA TCT AAA       1056
Ser Val Ala Ser Lys Asp Ala Lys Gln Lys Leu Val Ala Gln Ser Lys
            340             345             350

TAAACCAATA ATTAATGGGG ATGTGCATGC TAGCTAGCAT GTAAAGGCAA ATTAGGTTGT      1116

AAACCTCTTT GCTAGCTATA TTGAAATAAA CCAAAGGAGT AGTGTGCATG TCAATTCGAT      1176

TTTGCCATGT ACCTCTTGGA ATATTATGTA ATAATTATTT GAATCTCTTT AAGGTACTTA      1236

                                                                       1244

ATTAATCA
```

(2) INFORMATION FOR SEQ ID NO: 2:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 4700 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (iii) HYPOTHETICAL: NO

    (iv) ANTI-SENSE: NO

    (ix) FEATURE:
        (A) NAME/KEY: promoter
        (B) LOCATION:1..1532

    (ix) FEATURE:
        (A) NAME/KEY: sig_peptide
        (B) LOCATION:1533..1610

    (ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION:1533..1751

    (ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION:2383..2574

    (ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION:3605..3769

    (ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION:4033..4515

    (ix) FEATURE:
        (A) NAME/KEY: intron
        (B) LOCATION:1752..2382

(ix) FEATURE:
            (A) NAME/KEY: intron
            (B) LOCATION:2575..3604

        (ix) FEATURE:
            (A) NAME/KEY: intron
            (B) LOCATION:3770..4032

        (ix) FEATURE:
            (A) NAME/KEY: CDS
            (B) LOCATION:1533..1751

        (ix) FEATURE:
            (A) NAME/KEY: CDS
            (B) LOCATION:2383..2574

        (ix) FEATURE:
            (A) NAME/KEY: CDS
            (B) LOCATION:3605..3769

        (ix) FEATURE:
            (A) NAME/KEY: CDS
            (B) LOCATION:4033..4516

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

TAGATAAAAA AATGGGATAT AATTTTTCTC AGATGTTGTT TATACTGTTT TTTTAATCAG   60

AATTAAAATT CCTCTTTAAT TATCGACATA ATTTTTTTTG GTGAATATTA TCGACATAAT   120

TATTTAATAC AAATTTTTAT TGTACATAGA AGTGATACTT CAATTTTAAT ATTGGAGAAC   180

AGTACGAAAA CATAAAAAAA CTGTTATTAG AAGAAAAAAA TATATGGAAA AGGTTAGCTA   240

CATATATTAG CTAAATTAGT TGTTCTAATT GGCTATATAA ACCCTATTGT ACTCTTTGTA   300

ATCTCACCTT TTTCATTTAA ATACATTTCT ACTTTTTAAG TTCTATATTT TCTCTCAATT   360

TTCTTCGATA AACCATGAAA TTTAACATGG TATATCAGCG ATACCACCCA CTTTGAAAGC   420

CATGTATGGC TAGTATGGGC AGCCAAAATT TGCCCTGGTT CAAGCAAAGC AAGTGTTTAT   480

ATAGATGTGA CTTTTGTTGA GGAACTCATG CCAATGGTAC TGATTGTGAA ACTGAGAAAA   540

CTAATTTGGA GAATTTGAAT TATGATCATT AAATACTCCT CTCCTGACTA CCTTCGTCCC   600

TCAAATTTGT ACCATCATTA TTTCCCAAAA ATTTGATTAC AATGCACTAA TTAATGAATG   660

TTTCTTACAT TATCATATTA TCATATCTGA CATTTGTTT TTACTTTTTA TAATAATTAT   720

TTTAAAAGT CATACATGCA AATAATTTTT TAATAGTTTA CAGTTAAATT TTTACAGTAA   780

AAATGCATGA AAATTAAACT TTATTTTTCC AAGTCATCAT TTAGTCAAAT CCCAAAACAA   840

TGATTATTTT TTGCAAATGA ATGTTTATTG AACATTTAAA TGTAGCCTAA TTAATTCTGG   900

TTATGGTGTC AATGTTCCAA AACCTAATGC AAGATCTTAG CAAGTACATA CATAGATCTA   960

ATTTTAAACT TATCTTTACG CAAGAGATAT AAAGATTATA CATCTAGTTT TAAACATTAA   1020

CTTTTGTTTT TGTGTTAAAA AACAGTAACA TTTTCTTAAT TTTGTAGAGT GACGTGCTCC   1080

AACCATATTA ACGAAGATTT TAATTGGTAT TCAAGTTCAT GAACTTAGTA AATAAGTTTT   1140

GGTCTTCAGT TTTCAATTTT CATTACAACA TTTATGTAAA ATATCAACGT TTTCTGAAAT   1200

TTGTTGCTTG TGTGCTCCAA CCACATTTAA GAGATTATAG AAATTAATTT TCAAGAAGAT   1260

AATGATTCCT ACTCTTGCTG GCCCTACCAT AGTACAATAA ATCCACTCAT AAATCAACAA   1320

GTCGTCGTCA TAGGCAATTG GGCATCATAT CATAAACAAT ACGTACGTGA TATTATCTAG   1380

TGTCTCTCAG TTTACTTTAT GAGAAATTAT TTTTCTTTAA AAAAGTTAA TTAATAAAAA   1440

CATTTGCGAT ACCGTGAGTT ACAAGAAATC CGCCGAATTC ATCTCTATAA ATAAAGGAT   1500

CTATATGAGA GGTAAAATCA TATTAACTCA AA ATG GGT TCC ATG CGT CTA TTA   1553
                                     Met Gly Ser Met Arg Leu Leu
                                      1   5

GTA GTG GCA TTG TTG TGT GCA TTT GCT ATG CAT GCA GGT TTT TCA GTC   1601
Val Val Ala Leu Leu Cys Ala Phe Ala Met His Ala Gly Phe Ser Val
         10   15   20

TCT TAT GCT CAG CTT ACT CCT ACG TTC TAC AGA GAA ACA TGT CCA AAT   1649
Ser Tyr Ala Gln Leu Thr Pro Thr Phe Tyr Arg Glu Thr Cys Pro Asn
     25   30   35

CTG TTC CCT ATT GTG TTT GGA GTA ATC TTC GAT GCT TCT TTC ACC GAT   1697
Leu Phe Pro Ile Val Phe Gly Val Ile Phe Asp Ala Ser Phe Thr Asp
40   45   50   55

CCC CGA ATC GGG GCC AGT CTC ATG AGG CTT CAT TTT CAT GAT TGC TTT   1745
Pro Arg Ile Gly Ala Ser Leu Met Arg Leu His Phe His Asp Cys Phe
             60   65   70

GTT CAA GTACGTACTT TTTTTTTTCC TTCCAAAATG CCCTGCATAT TTAACAAGAT   1801
Val Gln

TGCTTTGTTC ACCTAGAAAA ATGTGTTTTT TTCAACGATC TTACGTACGT TTGTTTGGTT   1861

TGAAAAATAA ATCAGAAAGA GATCAAGAAA ATAGCTAGAA AGAAAGCAAC GTTTTTTTAA   1921

AAGGTATTTA GTGTGAGAAA AATATTAAAA CTGAAGAGAA AGAAATTAAA TAAGCTTTTC   1981

TTGAATGATA TTTACATGTC TTATTAACTT AAAGTCACCT TTTTTCTTTA AGTTGTGCTT   2041

GAAGAAAAAA GATGTCTTTC AGTTTAGTTT TGATTAATGC TAATTATATT TTTAATTAAT   2101

TAATTAATAC TATATATCTA TTTACCATAT TAATTATTAC TATATTTCAT GATGACAACA   2161

GACAAGTATT CTAAAGAGGT ATCGGTAGAT GATTAATTTT TTTATAAAAA AATCTTTTGC   2221

GTGTATAGAT ATTCTTTTAT AATTGGTGCA GAAACTTGTA ATGCTAATTG CAATTAATCT   2281

TACATTGATT AACTAATAGC TATAATCAAT ATTTAGGTTA GGTATAGGAG ACAAATCAAG   2341

TGATCTGAAC AAATTAAGTT GTTATATTTG CATTGTGACA G GGT TGT GAT GGA   2394
                                               Gly Cys Asp Gly
                                                      75

TCA GTT TTG CTG AAC AAC ACT GAT ACA ATA GAA AGC GAG CAA GAT GCA   2442
Ser Val Leu Leu Asn Asn Thr Asp Thr Ile Glu Ser Glu Gln Asp Ala
         80   85   90

```
CTT CCA AAT ATC AAC TCA ATA AGA GGA TTG GAC GTT GTC AAT GAC ATC        2490
Leu Pro Asn Ile Asn Ser Ile Arg Gly Leu Asp Val Val Asn Asp Ile
     95              100             105

AAG ACA GCG GTG GAA AAT AGT TGT CCA GAC ACA GTT TCT TGT GCT GAT        2538
Lys Thr Ala Val Glu Asn Ser Cys Pro Asp Thr Val Ser Cys Ala Asp
110             115             120             125

ATT CTT GCT ATT GCA GCT GAA ATA GCT TCT GTT CTG GTAATTAATA            2584
Ile Leu Ala Ile Ala Ala Glu Ile Ala Ser Val Leu
             130             135

ACTCCTAATT AATTCCCAAC CATTAAAAAG TTGCATGATT GGATTCAAAA TTCTATGGTA      2644

TTGGGGTTCT GATATAAATT TGTAATTAAA TTGCACTAAA AAAAATTATC ATATACTTTT      2704

AATAAAAAAA ATTTATCTAA TTTAATTTAT TATTAAAACT ATTTTTAAAA TTCAATCCTA      2764

ACTCTTTTTT AATCGGAGCA TGTAAGCTGG CACCCACCGT ATATCGTTGG AAGATGCTAT      2824

AAAACCATTT AATTAATGGA TGGAATCAGT CAAAACATTT AATTCAAAAT ACTCTTAATT      2884

GTGATTAGTA ATCATGTTCG GGCAAGTTAC GTTGTGTATA ATTAATTTGA CTTAATCAGA      2944

TAAAAAAACA AATGGACGCA AGCCGGTTGG TATAGATATC ACTGGCCTGT AGAATATGTG      3004

GTTTTTCACG TTTAAATAAA AGCTAGCTAC TATATTATAT TTAGTCTTTT TTTTTCTTAA      3064

ACCCATTTAA CGTGATTTAT TGACTGTGAA ACATGTTTCC ACACACAGGC TTAGAAACTC      3124

CTCGCAACTA ACATCTCCAA AATTTGACTA TTTATTTATG AAGATAATTC ATCTATGATG      3184

TTCAACTCTA TTATATATAT GTATCATCGC AGTATTAAGA ATTATAATAG TCAAATATAG      3244

AAGTATATCG GGTAAATGTA GTTGCATGTG CGACCTGTTT CGTGTAAAAT GCTTATTCTA      3304

TATAGCTTTT TTTATTGGAA AATAACGATG AACTAAAAAC GAAAGGGTAT CATATAGTTT      3364

GACTTTTATG TTAGAGAGAG ACATCTTAAT TTGGTCATAT GTTAAATAAT TAATTACAAT      3424

GCATACACAA ATATTTATGC CATATCTAAA AAATGATAAA ATATCATAGG TATACTCAAC      3484

TATATGATAT CCCCATAACA GAAATTGTAC TTTTCTTCAG GCAATGAACT TAACATTTCT      3544

GTTTGCTAAA AACAAACATC CACTTAAAGT GGTTCAACAT ATTTATGTAA TAATTTACAG      3604

GGA GGA GGT CCA GGA TGG CCA GTT CCA TTA GGA AGA AGG GAC AGC TTA       3652
Gly Gly Gly Pro Gly Trp Pro Val Pro Leu Gly Arg Arg Asp Ser Leu
         140             145             150

ACA GCA AAC CGA ACC CTT GCA AAT CAA AAC CTT CCA GCA CCT TTC TTC       3700
Thr Ala Asn Arg Thr Leu Ala Asn Gln Asn Leu Pro Ala Pro Phe Phe
     155             160             165

AAC CTC ACT CAA CTT AAA GCT TCC TTT GCT GTT CAA GGT CTC AAC ACC       3748
Asn Leu Thr Gln Leu Lys Ala Ser Phe Ala Val Gln Gly Leu Asn Thr
170             175             180             185

CTT GAT TTA GTT ACA CTC TCA GGTATACATA ATCAATTTTT TATTTGCTAT          3799
Leu Asp Leu Val Thr Leu Ser
             190

TAGCTAGCAA TAAAAAGTCT CTGATACAGA CATATTTAGA TAAATTAATT TCTCCATAAA     3859
```

CATTTATAAT AAAATTATCA ATTTATGTAC TTAAAAATTA TGGATTGAAG CTCTTTTCAT   3919

CCAACTTTTA CTAAAGTTAA GGTGCATATA ATATAAAATA AACTATCTCT TGTTTCTTAT   3979

AAAAAGATTG AAGATAAGTT AAAGTCTACT TATAAATCAT TAATATATGT ATA GGT   4035
                                                             Gly

GGT CAT ACG TTT GGA AGA GCT CGG TGC AGT ACA TTC ATA AAC CGA TTA   4083
Gly His Thr Phe Gly Arg Ala Arg Cys Ser Thr Phe Ile Asn Arg Leu
    195   200   205

TAC AAC TTC AGC AAC ACT GGA AAC CCT GAT CCA ACT CTG AAC ACA ACA   4131
Tyr Asn Phe Ser Asn Thr Gly Asn Pro Asp Pro Thr Leu Asn Thr Thr
210   215   220   225

TAC TTA GAA GTA TTG CGT GCA AGA TGC CCC CAG AAT GCA ACT GGG GAT   4179
Tyr Leu Glu Val Leu Arg Ala Arg Cys Pro Gln Asn Ala Thr Gly Asp
            230   235   240

AAC CTC ACC AAT TTG GAC CTG AGC ACA CCT GAT CAA TTT GAC AAC AGA   4227
Asn Leu Thr Asn Leu Asp Leu Ser Thr Pro Asp Gln Phe Asp Asn Arg
            245   250   255

TAC TAC TCC AAT CTT CTG CAG CTC AAT GGC TTA CTT CAG AGT GAC CAA   4275
Tyr Tyr Ser Asn Leu Leu Gln Leu Asn Gly Leu Leu Gln Ser Asp Gln
            260   265   270

GAA CTT TTC TCC ACT CCT GGT GCT GAT ACC ATT CCC ATT GTC AAT AGC   4323
Glu Leu Phe Ser Thr Pro Gly Ala Asp Thr Ile Pro Ile Val Asn Ser
            275   280   285

TTC AGC AGT AAC CAG AAT ACT TTC TTT TCC AAC TTT AGA GTT TCA ATG   4371
Phe Ser Ser Asn Gln Asn Thr Phe Phe Ser Asn Phe Arg Val Ser Met
290   295   300   305

ATA AAA ATG GGT AAT ATT GGA GTG CTG ACT GGG GAT GAA GGA GAA ATT   4419
Ile Lys Met Gly Asn Ile Gly Val Leu Thr Gly Asp Glu Gly Glu Ile
            310   315   320

CGC TTG CAA TGT AAT TTT GTG AAT GGA GAC TCG TTT GGA TTA GCT AGT   4467
Arg Leu Gln Cys Asn Phe Val Asn Gly Asp Ser Phe Gly Leu Ala Ser
            325   330   335

GTG GCG TCC AAA GAT GCT AAA CAA AAG CTT GTT GCT CAA TCT AAA TAA   4515
Val Ala Ser Lys Asp Ala Lys Gln Lys Leu Val Ala Gln Ser Lys *
            340   345   350

ACCAATAATT AATGGGGATG TGCATGCTAG CTAGCATGTA AAGGCAAATT AGGTTGTAAA   4575

CCTCTTTGCT AGCTATATTG AAATAAACCA AAGGAGTAGT GTGCATGTCA ATTCGATTTT   4635

GCCATGTACC TCTTGGAATA TTATGTAATA ATTATTTGAA TCTCTTTAAG GTACTTAATT   4695

                                                                    4700

AATCA

    (2) INFORMATION FOR SEQ ID NO: 3:
        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 17 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TTYCAYGAYT GYTTYGT   17

(2) INFORMATION FOR SEQ ID NO: 4:

          (i) SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 20 base pairs
                (B) TYPE: nucleic acid
                (C) STRANDEDNESS: single
                (D) TOPOLOGY: linear

          (ii) MOLECULE TYPE: DNA

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:   17

CTTCCAAATA TCAACTCAAT

(2) INFORMATION FOR SEQ ID NO: 5:

          (i) SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 20 base pairs
                (B) TYPE: nucleic acid
                (C) STRANDEDNESS: single
                (D) TOPOLOGY: linear

          (ii) MOLECULE TYPE: DNA

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:   20

TAAAGTTGGA AAAGAAAGTA

(2) INFORMATION FOR SEQ ID NO: 6:

          (i) SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 20 base pairs
                (B) TYPE: nucleic acid
                (C) STRANDEDNESS: single
                (D) TOPOLOGY: linear

          (ii) MOLECULE TYPE: DNA

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:   20

ATGCATGCAG GTTTTTCAGT

(2) INFORMATION FOR SEQ ID NO: 7:

          (i) SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 20 base pairs
                (B) TYPE: nucleic acid
                (C) STRANDEDNESS: single
                (D) TOPOLOGY: linear

          (ii) MOLECULE TYPE: DNA

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:   20

TTGCTCGCTT TCTATTGTAT

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 20 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
TCTTCGATGC TTCTTTCACC                                                    20
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 21 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
CATAAACAAT ACGTACGTGA T                                                  21
```

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 1031 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
TTTCATGATT GCTTTGTTCA AGGTTGTGAT GGATCAGTTT TACTGAACAA CACTGATACA   60
ATAGAAAGCG AGCAAGATGC ACTTCCAAAT ATCAACTCAA TAAGAGGATT GGACGTTGTC   120
AATGACATCA AGACAGCGGT GGAAAATAGT TGTCCAGACA CAGTTTCTTG TGCTGATATT   180
CTTGCTATTG CAGCTGAAAT AGCTTCTGTT GCTGGGAGGA GGTCAGGATG GCCAGTTCCA   240
TTAGGAAGAA GGGACAGCTT AACAGCAAAC CGAACCCTTG CAAATCAAAA CCTTCCAGCA   300
CCTTTCTTCA ACCTCACTCA ACTTAAAGCT TCCTTTGCTG TTCAAGGTCT CAACACCCTT   360
GATTTAGTTA CACTCTCAGG TGGTCATACG TCTGGAAGAG CTCGGTGCAG TACATTCATA   420
AACCGATTAT ACAACTTCAG CAACACTGGA CTGATCCACT TGGACACAAC ATACTTAGAA   480
GTATTGCGTG CAAGATGCCC CCAGAATGCA ACTGGGGATA ACCTCACCAA TTTGGACCTG   540
AGCACACCTG ATCAATTTGA CAACAGATAC TACTCCAATC TTCTGCAGCT CAATGGCTTA   600
CTTCAGAGTG ACCAAGAACG TTTCTCCACT CCTGGTGCTG ATACCATTCC ATTGTCAATA   660
GCTTCAGCGA ACCAGAATAC TTTCTTTTCC AACTTTAGAG TTTCAATGAT AAAAATGGGT   720
AATATTGGAG TGCTGACTGG GGATGAAGGA GAAATTCGCT TGCAATGTAA TTTTGTGAAT   780
```

GGAGACTCGT TTGGATTAGC TAGTGTGGCG TCCAAAGATG CTAAACAAAA GCTTGTTGCT 840

CAATCTAAAT AAACCAATAA TTAATGGGGA TGTCGATGCT AGCTACGATG TAAAGGCAAA 900

TTAGGTTGAA ACCTCTTTGC TAGCTATATT GAAATAAACC AAAGGAGTAG TGTCGATGTC 960

AATTCGATTT TGCCATGTAC CTCTTGGAAT ATTATGTAAT AATTATTTGA ATCTCAAAAA 1020

                                                                 1031

AAAAAAAAAA A

(2) INFORMATION FOR SEQ ID NO: 11:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1200 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGCAAACAAT GAACTCCCTT CGTGCTGTAG CAATAGCTTT GTGCTGTATT GTGGTTGTGC 60

TTGGAGGGTT ACCCTTCTCT TCAAATGCGC AACTTGATCC ATCCTTTTAC AGGAACACTT 120

GTCCAAATGT TAGTTCCATT GTTCGTGAAG TCATAAGGAG TGTTTCTAAG AAAGATCCTC 180

GTATGCTTGC TAGTCTTGTC AGGCTTCACT TTCATGACTG TTTTGTTCAA GGTTGTGATG 240

CATCAGTTTT ACTAAACAAA ACTGATACCG TTGTGAGTGA ACAAGATGCT TTTCCAAACA 300

GAAACTCATT AAGAGGTTTG GATGTTGTGA ATCAAATCAA AACAGCTGTG GAAAAGGCTT 360

GTCCTAACAC AGTTTCTTGT GCTGATATTC TTGCTCTTTC TGCTGAATTA TCATCTACAC 420

TGGCAGATGG TCCTGACTGG AAGGTTCCTT TAGGAAGAAG AGATGGTTTA ACGGCAAACC 480

AGTTACTTGC TAATCAAAAT CTTCCAGCTC CTTTCAATAC TACTGATCAA CTTAAAGCTG 540

CATTTGCTGC TCAAGGTCTC GATACTACTG ATCTGGTTGC ACTCTCCGGT GCTCATACAT 600

TTGGAAGAGC TCATTGCTCT TTATTTGTTA GCCGATTGTA CAACTTCAGC GGTACGGGAA 660

GTCCCGATCC AACTCTTAAC ACAACTTACT TACAACAATT GCGCACAATA TGTCCCAATG 720

GTGGACCTGG CACGAACCTT ACCAATTTCG ATCCAACGAC TCCTGATAAA TTTGACAAGA 780

ACTATTACTC TAATCTTCAA GTGAAAAAAG GTTTGCTTCA AAGTGATCAA GAGTTGTTCT 840

CAACATCTGG TTCAGATACC ATTAGCATTG TCAACAAATT CGCAACCGAT CAAAAAGCTT 900

TTTTTGAGAG CTTTAGGGCT GCTATGATCA AAATGGGAAA TATTGGTGTG TTAACCGGGA 960

ACCAAGGAGA GATTAGAAAA CAATGCAACT TTGTTAATTC AAAATCAGCA GAACTTGGTC 1020

TTATCAATGT TGCCTCAGCA GATTCATCTG AGGAGGGTAT GGTTAGCTCA ATGTAAATGT 1080

AGTGATTGGA AGCAACTAAT AAATTAAGAA GCTATAACTA TGCACATTCA TGGTATGTGT 1140

GAGATAGTTA TTAGATGCTT TGTGAGCAAA AATCTTTTGG ATTTCATTTG AAGTGTTTCT 1200

(2) INFORMATION FOR SEQ ID NO: 12:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1200 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```
GCTCTTCAAA ACAATGAACT CCTTAGCAAC TTCTATGTGG TGTGTTGTGC TTTTAGTTGT 60

GCTTGGAGGA CTACCCTTTT CCTCAGATGC ACAACTTAGT CCCACTTTTT ACAGCAAAAC 120

GTGTCCAACT GTTAGTTCCA TTGTTAGCAA TGTCTTAACA AACGTTTCTA AGACAGATCC 180

TCGCATGCTT GCTAGTCTCG TCAGGCTTCA CTTTCATGAC TGTTTTGTTC TGGGATGTGA 240

TGCCTCAGTT TTGCTGAACA ATACTGCTAC AATCGTAAGC GAACAACAAG CTTTTCCAAA 300

TAACAACTCT CTAAGAGGTT TGGATGTTGT GAATCAGATC AAACTGGCTG TAGAAGTGCC 360

TTGTCCTAAC ACAGTTTCTT GTGCTGATAT TCTTGCACTT GCTGCTCAAG CATCCTCTGT 420

TCTGGCACAA GGTCCTAGTT GGACGGTTCC TTTAGGAAGA AGGGATGGTT TAACCGCAAA 480

CCGAACACTT GCAAATCAAA ATCTTCCGGC TCCATTCAAT TCCTTGGATC AACTTAAAGC 540

TGCATTTACT GCTCAAGGCC TCAATACTAC TGATCTAGTT GCACTCTCGG GTGCTCATAC 600

ATTTGGAAGA GCTCATTGCG CACAATTTGT TAGTCGATTG TACAACTTCA GCAGTACTGG 660

AAGTCCCGAT CCAACTCTTA ACACAACTTA CTTACAACAA CTGCGCACAA TATGTCCCAA 720

TGGTGGACCT GGCACAAACC TTACCAATTT CGATCCAACG ACTCCTGATA AATTTGACAA 780

GAACTATTAC TCCAATCTTC AAGTGAAAAA GGGTTTGCTC CAAAGTGATC AAGAGTTGTT 840

CTCAACTTCT GGTGCAGATA CCATTAGCAT TGTCAACAAA TTCAGCACCG ATCAAAATGC 900

TTTCTTTGAG AGCTTTAAGG CTGCAATGAT TAAAATGGGC AATATTGGTG TGCTAACAGG 960

GACAAAAGGA GAGATTAGAA AACAATGCAA CTTTGTGAAC TTTGTGAACT CAAATTCTGC 1020

AGAACTAGAT TTAGCCACCA TAGCATCCAT AGTAGAATCA TTAGAGGATG GTATTGCTAG 1080

TGTAATATAA ATAAATTAGC GTAAATGCAC TTATTGAAAT CTTGTGACTA GATGCCACTA 1140

ATAAATAAGT TATAACTAGG CACATTTCAT GTCACTTGAA ATTTCATGCC TGTATATGAG 1200
```

(2) INFORMATION FOR SEQ ID NO: 13:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1200 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

```
CTCCTTAGCA ACTTCTATGT GGTGTGTTGT GCTTTTAGTT GTGCTTGGAG GACTACCCTT   60
TTCCTCAGAT GCACAACTTA GTCCCACTTT TTACAGCAAA ACGTGTCCAA CTGTTAGTTC  120
CATTGTTAGC AATGTCTTAA CAAACGTTTC TAAGACAGAT CCTCGCATGC TTGCTAGTCT  180
CGTCAGGCTT CACTTTCATG ACTGTTTTGT TCTGGGATGT GATGCCTCAG TTTTGCTGAA  240
CAATACTGCT ACAATCGTAA GCGAACAACA AGCTTTTCCA AATAACAACT CTCTAAGGGG  300
TTTGGATGTT GTGAATCAGA TCAAAACTGC TGTAGAAAGT GCTTGTCCTA ACACAGTTTC  360
TTGTGCTGAT ATTCTTGCAC TTGCTCAAGC ATCCTCTGTT CTGGCACAAG GTCCTAGTTG  420
GACGGTTCCT TTAGGAAGAA GGGATGGTTT AACCGCAAAC CGAACACTTG CAAATCAAAA  480
TCTTCCGGCT CCATTCAATT CCTTGGATCA CCTTAAACTG CATTTGACTG CTCAAGGCCT  540
CATTACTCCT GTTCTAGTTG CCCTCTCGGG TGCTCATACA TTTGGAAGAG CTCATTGCGC  600
ACAATTTGTT AGTCGATTGT ACAACTTCAG CAGTACTGGA AGTCCCGATC CAACTCTTAA  660
CACAACTTAC TTACAACAAC TGCGCACAAT ATGTCCCAAT GGTGGACCTG GCACAAACCT  720
TACCAATTTC GATCCAACGA CTCCTGATAA ATTTGACAAG AACTATTACT CCAATCTTCA  780
AGTGAAAAAG GGTTTGCTCC AAAGTGATCA AGAGTTGTTC TCAACTTCTG GTGCAGATAC  840
CATTAGCATT GTCGACAAAT TCAGCACCGA TCAAAATGCT TTCTTTGAGA GCTTTAAGGC  900
TGCAATGATT AAAATGGGCA ATATTGGTGT GCTAACAGGG ACAAAAGGAG AGATTAGAAA  960
ACAATGCAAC TTTGTGAACT CAAATTCTGC AGAACTAGAT TTAGCCACCA TAGCATCCAT 1020
AGTAGAATCA TTAGAGGATG GAATTGCTAG TGTAATATAA ATAAATTAGC GAAAATGCAC 1080
TTATTGAAAT CTTGTGACTA GATCCCACTA ATAAATAAGT TATAACTAGG CACATTTCAT 1140
GTCACTTGAA ATCCTATGCC TTGTATATTA GAGGACGTGT TCTTCTTGGT ATTATACTAT 1200
```

(2) INFORMATION FOR SEQ ID NO: 14:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1200 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
AATGCTTGGT CTAAGTGCAA CAGCTTTTTG CTGTATGGTG TTTGTGCTAA TTGGAGGAGT   60
ACCCTTTTCA AATGCACAAC TAGATCCTTC ATTTTACAAC AGTACATGTT CTAATCTTGA  120
TTCAATCGTA CGTGGTGTGC TCACAAATGT TTCACAATCT GATCCCAGAA TGCTTGGTAG  180
```

```
TCTCATCAGG CTACATTTTC ATGACTGTTT TGTTCAAGGT TGCGATGCCT CGATTTTGCT 240

GAACGATACG GCTACAATAG TGAGCGAGCA AAGTGCACCA CCAAATAACA ACTCCATAAG 300

AGGTTTGGAT GTGATAAACC AGATCAAAAC AGCGGTGGAA AATGCTTGTC CTAACACAGT 360

TTCTTGTGCT GATATTCTTG CTCTTTCTGC TGAAATATCA TCTGATCTGG CAAATGGTCC 420

TACTTGGCAA GTTCCATTAG GAAGAAGGGA TAGTTTGACA GCAAATAATT CCCTTGCAGC 480

TCAAAATCTT CCTGCCCCCA CTTTCAACCT TACTCGACTA AAATCTAACT TTGATAATCA 540

AAACCTCAGT ACTACTGATC TAGTTGCACT CTCAGGTGGC CATACAATTG GAAGAGGTCA 600

ATGCAGATTT TTCGTTGATC GATTATACAA TTTCAGCAAC ACTGGAAACC CCGATTCAAC 660

TCTTAACACG ACCTATTTAC AAACATTGCA AGCAATATGT CCCAATGGTG GACCTGGTAC 720

AAACCTAACC GATTTGGACC CAACCACACC AGATACATTT GACTCCAACT ACTACTCCAA 780

TCTCCAAGTT GGAAAGGGCT TGTTTCAGAG TGACCAAGAG CTTTTTTCCA GAAATGGTTC 840

TGACACTATT TCTATTGTCA ATAGTTTCGC CAATAATCAA ACTCTCTTCT TTGAAAATTT 900

TGTAGCCTCA ATGATAAAAA TGGGTAATAT TGGAGTTTTA ACTGGATCTC AAGGTGAAAT 960

TAGAACACAG TGTAATGCTG TGAATGGGAA TTCTTCTGGA TTGGCTACTG TAGTCACCAA 1020

AGAATCATCA GAAGATGGAA TGGCTAGCTC ATTCTAAATA TAAGCTTGGA AAATATTGAA 1080

GAGGTTCTAT AATTTTGTGC ATACATATAT GGTATGTGCA TGTGGTGTAT TATGTTTTTG 1140

TTATGTTCTT CAAGTTGATC AGGGACTGTA GAAGCTCCCT AATAATATTT GTGTCAAAGT 1200
```

(2) INFORMATION FOR SEQ ID NO: 15:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 283
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: prt

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
FHDCFVQGCD GSVLLNNTDT IESEQDALPN INSIRGLDVV NDIKTAVENS CPDTVSCADI 60

LAIAAEIASV AGRRSGWPVP LGRRDSLTAN RTLANQNLPA PFFNLTQLKA SFAVQGLNTL 120

DLVTLSGGHT SGRARCSTFI NRLYNFSNTG LIHLDTTYLE VLRARCPQNA TGDNLTNLDL 180

STPDQFDNRY YSNLLQLNGL LQSDQERFST PGADTIPLSI ASANQNTFFS NFRVSMIKMG 240
                                                                        283
NIGVLTGDEG EIRLQCNFVN GDSFGLASVA SKDAKQKLVA QSK
```

(2) INFORMATION FOR SEQ ID NO: 16:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 355

```
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: PRT


    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:


MNSLRAVAIA LCCIVVVLGG LPFSSNAQLD PSFYRNTCPN VSSIVREVIR SVSKKDPRML 60

ASLVRLHFHD CFVQGCDASV LLNKTDTVVS EQDAFPNRNS LRGLDVVNQI KTAVEKACPN 120

TVSCADILAL SAELSSTLAD GPDWKVPLGR RDGLTANQLL ANQNLPAPFN TTDQLKAAFA 180

AQGLDTTDLV ALSGAHTFGR AHCSLFVSRL YNFSGTGSPD PTLNTTYLQQ LRTICPNGGP 240

GTNLTNFDPT TPDKFDKNYY SNLQVKKGLL QSDQELFSTS GSDTISIVNK FATDQKAFFE 300

SFRAAMIKMG NIGVLTGNQG EIRKQCNFVN SKSAELGLIN VASADSSEEG MVSSM      355


(2) INFORMATION FOR SEQ ID NO: 17:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 358
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: PRT


    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

MNSLATSMWC VVLLVVLGGL PFSSDAQLSP TFYSKTCPTV SSIVSNVLTN VSKTDPRMLA 60

SLVRLHFHDC FVLGCDASVL LNNTATIVSE QQAFPNNNSL RGLDVVNQIK LAVEVPCPNT 120

VSCADILALA AQASSVLAQG PSWTVPLGRR DGLTANRTLA NQNLPAPFNS LDQLKAAFTA 180

QGLNTTDLVA LSGAHTFGRA HCAQFVSRLY NFSSTGSPDP TLNTTYLQQL RTICPNGGPG 240

TNLTNFDPTT PDKFDKNYYS NLQVKKGLLQ SDQELFSTSG ADTISIVNKF STDQNAFFES 300

FKAAMIKMGN IGVLTGTKGE IRKQCNFVNF VNSNSAELDL ATIASIVESL EDGIASVI   358


(2) INFORMATION FOR SEQ ID NO: 18:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 347
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: PRT


    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:
```

```
MWCVVLLVVL GGLPFSSDAQ LSPTFYSKTC PTVSSIVSNV LTNVSKTDPR MLASLVRLHF 60

HDCFVLGCDA SVLLNNTATI VSEQQAFPNN NSLRGLDVVN QIKTAVESAC PNTVSCADIL 120

ALAQASSVLA QGPSWTVPLG RRDGLTANRT LANQNLPAPF NSLDHLKLHL TAQGLITPVL 180

VALSGAHTFG RAHCAQFVSR LYNFSSTGSP DPTLNTTYLQ QLRTICPNGG PGTNLTNFDP 240

TTPDKFDKNY YSNLQVKFGL LQSDQELFST SGADTISIVD KFSTDQNAFF ESFKAAMIKM 300

GNIGVLTGTK GEIRKQCNFV NSNSAELDLA TIASIVESLE DGIASVI          347
```

(2) INFORMATION FOR SEQ ID NO: 19:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 351
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: PRT

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

```
MLGLSATAFC CMVFVLIGGV PFSNAQLDPS FYNSTCSNLD SIVRGVLTNV SQSDPRMLGS 60

LIRLHFHDCF VQGCDASILL NDTATIVSEQ SAPPNNNSIR GLDVINQIKT AVENACPNTV 120

SCADILALSA EISSDLANGP TWQVPLGRRD SLTANNSLAA QNLPAPTFNL TRLKSNFDNQ 180

NLSTTDLVAL SGGHTIGRGQ CRFFVDRLYN FSNTGNPDST LNTTYLQTLQ AICPNGGPGT 240

NLTDLDPTTP DTFDSNYYSN LQVGKGLFQS DQELFSRNGS DTISIVNSFA NNQTLFFENF 300

VASMIKMGNI GVLTGSQGEI RTQCNAVNGN SSGLATVVTK ESSEDGMASS F          351
```